AD-A020 428

# PROBABILITIES OF EXCESSIVE DEVIATIONS OF SIMPLE LINEAR RANK STATISTICS

Robert P. Clickner, et al

Florida State University

Prepared for:

Army Research Office

February 1975

# The Florida State University

# Department of Statistics

Tallahassee, Florida 32306

Probabilities of Excessive Deviations
of Simple Linear Rank Statistics

by

Robert P. Clickner[1,2] and Jayaram Sethuraman[3]
The Florida State University

FSU Technical Report No. M311

USARO-D Technical Report No. 12

February, 1975

Abbreviated Title: Deviations of Linear Rank Statistics



[1,3]Research supported by the United States Army Research Office, Durham under Grant No. DAHC04-74-G-~~1000~~. 0188

[2]Research supported in part by the National Institutes of Health Grant No. 5-T01-GM00913 to the Florida State University.

The United States Government is authorized to reproduce and distribute reprints for government purposes.

[1]Now at Department of Statistics, Temple University, Philadelphia Pa., 19122.

[3]Now visiting professor at the University of Michigan.

Key words and phrases. Excessive deviations, convergence rates, simple linear rank statistics, asymptotic expansions.

AMS 1970 subject classifications 60F10, 62G99.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

| | |
|---|---|
| 1. REPORT NUMBER USARO-D 12    2. GOVT. ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and subtitle) Probabilities of Excessive Deviations of Simple Linear Rank Statistics | 5. TYPE OF REPORT & PERIOD COVERED Technical Report |
| | 6. PERFORMING ORG. REPORT NUMBER FSU Statistics Report M3[illegible] |
| 7. AUTHOR(s) Robert P. Clickner and Jayaram Sethuraman | 8. CONTRACT OR GRANT NUMBER(s) DAHC04 - 74 - G - 0188 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Florida State University Tallahassee, Florida 32306 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Box CM, Duke Station Durham, North Carolina 27706 | 12. REPORT DATE February 1975 |
| | 13. NUMBER OF PAGES |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | 15. SECURITY CLASS (of this report) Unclassified |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS

Excessive deviations, convergence rates, simple linear rank statistics, asymptotic expansions

20. ABSTRACT

Rates of convergence to 0 of probabilities of excessive deviations of simple linear rank statistics are obtained under the null hypothesis. These results fill the gap between the known results on the asymptotic normality and probabilities of large deviations of such statistics. They are also potentially useful in other applications like Bayes risk efficiency.

ia

## SUMMARY

### Probabilities of Excessive Deviations of Simple Linear Rank Statistics

Let $\{S_N\}$ be a sequence of r.v.'s whose asymptotic distribution is $N(0, \sigma_N^2)$ and let $\{x_N\}$ be a sequence of constants with $x_N \to \infty$. A (righthand) excessive deviation is an event of the form $\{S_N > x_N\sigma_N\}$. The asymptotic normality of $S_N$ tells us $P\{S_N > x_N\sigma_N\} \to 0$, $N \to \infty$, but not the rate of this convergence. These rates are needed for the evaluation of Bahadur ($x_N = x\sqrt{N}$) and Bayes Risk ($x_N = x\sqrt{\log N}$) efficiencies. When $S_N$ is a k-sample linear rank statistic (see Hàjek and Šidák (1967)), and $x_N^2 = o(N)$, we show $\log P\{S_N > x_N\sigma_N\} \sim - x_N^2/2$, assuming the null hypothesis. In the two-sample case, we establish also that $P\{S_N > x_N\sigma_N\} = o\{\exp[-x_N^2/2 + J(x_N, \lambda_N)]\}$, when $N\lambda_N$ $(0 < \lambda_N < 1)$ is the size of the first sample and J is a function whose behavior is analyzed for various $x_N$ and $\lambda_N$. For example, if $x_N^4 = o(N)$, then $J(x_N, \lambda_N) = o(1)$ as $N \to \infty$.

## 1. Introduction.

Simple linear rank statistics arise in a variety of situations, particularly in the problem of testing the equality of two or more distributions by non-parametric tests. While the asymptotic distributions of simple linear rank statistics have been studied extensively (see, e.g., Hàjek and Šidák (1967)), investigations concerning their large deviation properties have been made only recently (Stone (1967, 1968, 1969) and Woodworth (1970)). In this paper we study the rates of convergence to zero of null probabilities of excessive deviations of k-sample ($k \geq 2$) simple linear rank statistics. We begin in this section with discussions of the notions of an excessive deviation and of a simple linear rank statistic. The results are in Section 2 and the proofs are in Section 3.

The concept of an excessive deviation of a random variable will be discussed here in a general setting. Let $\{S_N, N > 1\}$ be a sequence of random variables with positive finite variances and

let $\{\sigma_N, N \geq 1\}$ be a sequence of positive constants such that

$$(1.1) \qquad \sup_x |P\{S_N \leq x\sigma_N\} - \Phi(x)| \to 0, \; N \to \infty,$$

where $\Phi(\cdot)$ is the standard normal distribution function.
Let $\{x_N, N \geq 1\}$ be a sequence of nonnegative real numbers.
We call $P_N(x_N) = P\{S_N > x_N\sigma_N\}$ the <u>probability of a</u> (right-hand) <u>deviation of size</u> $x_N$ <u>of</u> $S_N$. A trivial consequence of (1.1) is

$$(1.2) \qquad P_N(x_N) \sim \Phi(-x_N), \; N \to \infty,$$

provided $x_N = O(1)$. However, when $x_N \to \infty$, (1.2) is clearly no longer a direct consequence of (1.1), which leads one to ask: given $\{S_N\}$ and $\{x_N\}$, where $x_N \to \infty$ at some specified rate, does (1.2) hold? If not, what exactly is the asymptotic behavior of $P_N(x_N)$? To answer these questions, tools more refined than the central limit theorem are needed.

We will now introduce some terminology, due to Rubin and Sethuraman (1965a). If $x_N = O(1)$, the event $\{S_N > x_N\sigma_N\}$ is called an <u>ordinary deviation of</u> $S_N$, while, if $x_N \to \infty$, it is an <u>excessive deviation of</u> $S_N$. Two special cases of excessive deviations have

separate names, because of the applications they have found in statistics. The case where $x_N^2/\ell og\ N \to c^2$, $0 < c < \infty$ is a moderate deviation of $S_N$, which arises in the study of Bayes risk efficiency (see Rubin and Sethuraman (1965b) and Clickner (1972)). The most extensively studied excessive deviation is the large deviation, in which $x_N^2/N \to c^2$. This attention arises from the fact that probabilities of large deviations must be evaluated in order to compute Bahadur efficiencies (see Bahadur (1960)).

Much of the previous work on excessive deviations has been for sums of independent random variables. When $S_N$ is the mean of i.i.d. random variables with zero mean, unit variance and finite moment generating function, and $\sigma_N = N^{-\frac{1}{2}}$, Cramér (1938) has shown

$$(1.3) \qquad P_N(x_N) \sim \Phi(-x_N)\exp[x_N^3 N^{-\frac{1}{2}} \lambda(x_N N^{-\frac{1}{2}})],$$

where $x_N \to \infty$, $x_N^2 = O(N)$, and where $\lambda(z)$ is a function which admits a convergent power series expansion for small $|z|$. A corollary of (1.3) is that (1.2) holds if $x_N^6 = o(N)$.

Feller (1943) generalized Cramér's (1938) result to non-identically distributed random variables. Rubin and Sethuraman (1965a) considered only moderate deviations but were able to relax the requirement of a moment generating function. Other work on the problem has been done by Chernoff (1952), Linnik (1960, 1961, 1962), and others.

Only right-hand deviations $\{S_N > x_N\sigma_N\}$ have been discussed here. This practice will be continued throughout this paper. All results for right-hand excessive deviations have immediate extensions to left-hand deviations $\{S_N < -x_N\sigma_N\}$ and two-sided deviations $\{|S_N| > x_N\sigma_N\}$.

We will now define a simple linear rank statistic. Let $X_{11}, \ldots, X_{1n_1}, \ldots, X_{k1}, \ldots, X_{kn_k}$ be a sequence of independent and continuous random variables, where $k \geq 2$, $n_i \geq 1$, $i = 1, \ldots, k$ and $n_1 + \ldots + n_k = N$. Let $R_{11}, \ldots, R_{kn_k}$ be the ranks for the combined sample $X_{11}, \ldots, X_{kn_k}$. Consider the problem of testing the null hypothesis H: $X_{11}, \ldots, X_{kn_k}$ are identically distributed versus the k-sample alternative with density

$$(1.4) \qquad q(x_{11}, \ldots, x_{kn_k}) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} f(x_{ij} - \Delta_i),$$

where f is a known density and $\Delta_1, \ldots, \Delta_k$ are known constants.

A locally most powerful test of H versus q is based on the k sample simple linear rank statistic (see Hàjek and Šidák (1967) p. 69).

$$(1.5) \quad S_N = \sum_{i=1}^{k} \Delta_i \sum_{j=1}^{n_i} a_N(R_{ij})$$

where $a_N(1), \ldots, a_N(N)$ are a sequence of constants, called scores.

These simple linear rank statistics include many of the more common and important rank statistics. For example, if $a_N(i) = i/N+1$, $i=1, \ldots, N$, we obtain a k-sample extension of the classical Wilcoxon rank-sum statistic. When $a_N(i) = 1$ for $i > (N+1)/2$ and $-1$ for $i \leq (N+1)/2$, $S_N$ becomes a median test. With $a_N(i) = E[\Phi^{-1}(U_N^{(i)})]$, $i=1, \ldots, N$, where $U_N^{(1)} < \ldots < U_N^{(N)}$ are the order statistics for a sample of size N from the uniform (0,1) distribution, $S_N$ is the Fisher-Yates-Terry-Hoeffding normal scores statistic.

To the authors' knowledge, all previous work on excessive deviations of linear rank statistics has been for the case of large deviations only and under the null hypothesis H (Stone (1967, 1968, 1969) and Woodworth (1970)). In this paper we will be considering only excessive deviations that are not large (i.e. $x_N \to \infty$ but $x_N^2/N \to 0$) and only the null hypothesis H. Thus there is no overlap of the present paper with Stone's or Woodworth's work. In fact, our results fill in a gap between their work and the work of these many authors who have studied

the asymptotic normality of simple linear rank statistics (see Hàjek and Šidák (1967) for references.

## 2. Main Results.

We need some notation. Let $\lambda_{Ni} = n_i/N$, $i=1, \ldots, k$. Let $\mu_{\Delta N} = \lambda_{N1}\Delta_1 + \ldots + \lambda_{Nk}\Delta_k$ and

$$(2.1) \qquad \sigma^2_{\Delta N} = \sum_{i=1}^{k} \lambda_{Ni}(\Delta_i - \mu_{\Delta N})^2.$$

We shall assume

$$(2.2) \qquad 0 < \liminf_{N\to\infty} \lambda_{Ni} \leq \limsup_{N\to\infty} \lambda_{Ni} < 1, \; i = 1, \ldots, k,$$

and $\sigma^2_{\Delta N} > 0$ for all $N \geq k$. Further, let $\sum_{j=1}^{N} a_N(j) = 0$, and $\sum_{j=1}^{N} a^2_N(j) = N$.

Let $S_N$ be as defined in (1.5). Throughout the rest of this paper we will assume that the null hypothesis H holds. Then $ES_N = 0$ and var $S_N = N^2\sigma^2_{\Delta N}/(N-1)$. We define $P_N(x_N)$ to be

so we define $P_N(x_N)$ to be

$$(2.3) \qquad P_N(x_N) = P\{S_N \ge x_N\, \sigma_{\Delta N} N^{\frac{1}{2}}\}.$$

We can now state our main results.

Theorem 2.1. Let $S_N$ be of the form (1.5), assume the null hypothesis H obtains and define $P_N(x_N)$ by (2.3), where $x_N \to \infty$ and $x_N^2/N \to 0$. Assume

$$(2.4) \qquad x_N^2 \max_{1 \le j \le N} a_N^2(j) = O(N)$$

and

$$(2.5) \qquad \gamma_N^3 = \sum_{j=1}^{N} |a_N(j)|^3/N = o(N^{\frac{1}{2}}/x_N).$$

Then, as $N \to \infty$,

$$(2.6) \qquad \log P_N(x_N) \sim - x_N^2/2.$$

This crude estimate of $P_N(x_N)$ can be improved somewhat in the two-sample case ($k = 2$), provided (2.4) and (2.5) are strengthened a little. When $k = 2$, without loss of generality, we may set $\Delta_1 = 1$, $\Delta_2 = 0$, which gives (with $n=n_1$),

$$(2.7) \qquad S_N = \sum_{j=1}^{n} a_N(R_{1j}).$$

Letting $\lambda_N = \lambda_{N1}$ and $\overline{\lambda}_N = \lambda_{N2}$, we get $\mu_{\Delta N} = \lambda_N$ and $\sigma_{\Delta N}^2 = \lambda_N \overline{\lambda}_N$, so (2.3) becomes

$$(2.8)\qquad P_N(x_N) = P\{S_N \geq x_N(N\lambda_N\bar{\lambda}_N)^{\frac{1}{2}}\}.$$

We now state our additional results for the two-sample case. Observe that (2.9) includes both (2.4) and (2.5).

<u>Theorem 2.2</u>. Let the null hypothesis H obtain, let $S_N$ be of the form (2.7), and define $P_N(x_N)$ by (2.8), where $x_N \to \infty$, $x_N^2/N \to 0$. If

$$(2.9)\qquad \gamma_N^3 = o(N^{\frac{1}{2}}/x_N^3),$$

then, as $N \to \infty$,

$$(2.10)\qquad P_N(x_N) = o\{\exp[-x_N^2/2 + J(x_N,\lambda_N)]\},$$

where

$$(2.11)\qquad J(x_N,\lambda_N) = (x_N^2/2)(1 - \lambda_N\bar{\lambda}_N/p_Nq_N) + NI(\lambda_N,p_N),$$

where

$$(2.12)\qquad I(\lambda,p) = \lambda\log(\lambda/p) + \bar{\lambda}\log(\bar{\lambda}/q),$$

where $q = 1-p$, $q_N = 1-p_N$, and $p_N$ is the unique solution of the equation

$$(2.13)\qquad n = \sum_{j=1}^{N} \{1 + (q_N/p_N)\exp[-x_N a_N(j)(\lambda_N\bar{\lambda}_N/N)^{\frac{1}{2}}/p_Nq_N]\}^{-1}.$$

Theorems 2.1 and 2.2 are the best results obtainable with pre-

sently available methods. However, it would be desirable to improve on these results to obtain a better estimate of $P_N(x_N)$, one comparable to Cramér's (1938) result (1.3). To do this in the two-sample case a certain rate of convergence to normality of the normed distribution $\Phi_N(\cdot)$ of the mean of certain dependent random variables arising in the proof of Theorem 2.2 is required. This convergence rate has not been established. Let

$$(2.14) \qquad \Delta_N = \sup_x |\Phi_N(x) - \Phi(x)|,$$

where $\Phi_N(\cdot)$ is defined in (3.39). It follows from Hàjek (1964) that $\Delta_N \to 0$ as $N \to \infty$. His result is reproduced here as Lemma 3.7. The required convergence rate is $\Delta_N x_N = o(1)$ where $x_N \to \infty$ (See (3.46)). If it can be verified that this rate holds then the following estimate is obtainable:

$$(2.15) \qquad P_N(x_N) \sim \Phi(-x_N)\exp[J(x_N,\lambda_N)], \; N \to \infty.$$

To extend Theorem 2.2 to the case k > 2, a multivariate generalization of Hàjek's (1964) asymptotic normality result is required. Unfortunately, it too, is not yet available.

The conditions of Theorems 2.1 and 2.2 are stronger than those required to establish the asymptotic normality of simple linear rank statistics $S_N$ (See, e.g., Hàjek and Šidák (1967) pp. 193-195). This is to be expected since these theorems give

more detailed and precise information about the asymptotic behavior of $S_N$ than does a statement of asymptotic normality. Even so, it is natural to ask whether these stronger conditions are very restrictive, in terms of potential applications. Many simple linear rank statistics $S_N$ have scores of either of the forms

$$(2.16) \qquad a_N(i) = E\,\phi(U_N^{(i)})$$

or

$$(2.17) \qquad \hat{a}_N(i) = \phi(i/N+1),$$

where $\phi$ is a non-constant function on (0,1) and, further, $S_N$ is asymptotically normal if $\int_0^1 \phi^2(u)du < \infty$ (See, e.g., Hàjek and Šidák (1967) Chapter V). Clearly, if

$$(2.18) \qquad \int_0^1 |\phi(u)|^3 du < \infty,$$

then (2.9) holds for the scores $a_N$ and $\hat{a}_N$ and we may apply Theorems 2.1 and 2.2 to $S_N$. Some simple linear rank statistics that have scores of the form (2.16) or (2.17) where $\phi$ satisfies (2.18) are: median, Wilcoxon, Van der Waerden and Fisher-Yates-Terry-Hoeffding, all tests for location shift, and Capon, Klotz, Ansari-Bradley, quartile, and Savage, all tests for a shift in scale. See Clickner (1972) for details.

It is not immediately obvious from Theorem 2.2 as presented, exactly how fast $P_N(x_N)$ tends to zero when $x_N$ grows at a specified rate, say $x_N^2 = \log N$, or perhaps $x_N^8 = N^3$. It is necessary to analyze $J(x_N, \lambda_N)$ for various $x_N$ and $\lambda_N$ to see more clearly the behavior of $P_N(x_N)$. This is done in Corollaries 2.5 and 2.6 following the preliminary Lemmas 2.3 and 2.4.

Lemma 2.3. Let $p_N$ be the solution of equation (2.13). Let (2.9) hold. Then

$$(2.19) \quad p_N = \lambda_N + \sum_{k=1}^{\infty} c_k(\lambda_N)(x_N^2/N)^k,$$

where $\{c_k(\lambda),\ k \geq 1\}$ is a sequence of functions of $\lambda$ whose first two elements are

$$(2.20) \quad c_1(\lambda) = \tfrac{1}{2} - \lambda,$$

$$(2.21) \quad c_2(\lambda) = (\lambda - \tfrac{1}{2})(1 - 1/(2\lambda\bar{\lambda})).$$

Further, if $\lambda_N = \frac{1}{2} + O(N^{-1})$, then $p_N = \frac{1}{2} + O(N^{-1})$.

Lemma 2.4. Define $I(\lambda, p)$ by (2.12) for $0 < \lambda < 1$ and $0 < p < 1$. If $|\lambda - p| < \min(p, q)$, then

$$(2.22) \quad I(\lambda,p) = \tfrac{1}{2}\left(\frac{\lambda-p}{pq}\right)^2(2pq-\bar{\lambda}p^2+\lambda q^2) + \sum_{i=3}^{\infty} \frac{(\lambda-p)^i}{i}\left[\frac{\bar{\lambda}}{q^i} + \frac{\lambda}{(-p)^i}\right].$$

<u>Corollary 2.5</u>. In Theorem 2.2 assume $\lambda_N = \frac{1}{2} + O(1/N)$. Then $J(x_N,\lambda_N) = o(1)$.

<u>Corollary 2.6</u>. Under the conditions of Theorem 2.2,

$$(2.23)\quad J(x_N,\lambda_N) = \frac{x_N^4}{N}\sum_{i=0}^{\infty} d_i(\lambda_N)\left(\frac{x_N^2}{N}\right)^i ,$$

where $(d_i(\lambda),\ i \geq 0)$ is a sequence of functions of $\lambda$ whose first element is

$$(2.24)\quad d_0(\lambda) = -(1-2\lambda)^2/8\lambda\,\bar{\lambda}.$$

Further, if, for some integer $k = 1, 2, \ldots,$

$$(2.25)\quad \liminf(x_N^{2(k+1)}N^{-k}) > 0 \text{ and } x_N^{2(k+2)} = o(N^{k+1}),$$

then

$$(2.26)\quad J(x_N,\lambda_N) = o(1), \text{ if } k = 0$$

$$= \frac{x_N^4}{N}\sum_{i=0}^{k-1} d_i(\lambda_N)\left(\frac{x_N^2}{N}\right)^i + o(1), \text{ if } k > 0.$$

3. Proofs.

We begin the proofs with some preliminaries and four lemmas which will be used in proving both Theorems 2.1 and 2.2.

To simplify notation, we will often write $a_i$ for $a_N(i)$, $\lambda_i$ for $\lambda_{Ni}$, $\sigma_\Delta$ for $\sigma_{\Delta N}$, etc., suppressing the dependence on N. Let $W_{(1)} < \dots < W_{(N)}$ be the order statistics for the combined sample $X_{11}, \dots, X_{kn_k}$, and let

$$(3.1) \qquad C_{ij} = \begin{cases} 1 & \text{if } W_{(j)} \text{ is from the i-th sample,} \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, \dots, k$, $j = 1, \dots, N$. Clearly, $C_{i1} + \dots + C_{iN} = n_i$, $i = 1, \dots, k$. Then, by rearranging the sum,

$$(3.2) \qquad S_N = \sum_{j=1}^{N} a_j \sum_{i=1}^{k} \Delta_i C_{ij}.$$

Let $\Psi = \{(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)\}$ be a set consisting of k points with $z = (z_1, \dots, z_k)$ a typical element and let $Z^{(1)}, \dots, Z^{(N)}$ be a sequence of independent and identically distributed random vectors taking values in $\Psi$ with probabilities $p_1, \dots, p_k$, $p_i > 0$, $i = 1, \dots, k$, $p_1 + \dots p_k = 1$, that is,

$$(3.3) \qquad P\{Z^{(1)} = z\} = \prod_{i=1}^{k} p_i^{z_i} .$$

Let $Z(N) = Z^{(1)} + \ldots + Z^{(N)} = [Z_1(N), \ldots, Z_k(N)]$. Define a statistic $T_N$ by

(3.4) $$T_N = \sum_{j=1}^{N} a_j \sum_{i=1}^{k} \Delta_i Z_i^{(j)} ,$$

and let $t_N$ be a realization of $T_N$. The key to these proofs is the observation that

(3.5) $$P\{S_N = t_N\} = P\{T_N = t_N | Z(N) = \underline{n}\},$$

where $\underline{n} = (n_1, \ldots, n_k)$, and, further, (3.5) is an identity in the probabilities $p_1, \ldots, p_k$.

For each $h > 0$, define a new joint distribution for the random vectors $Z^{(1)}, \ldots, Z^{(N)}$ by

(3.6) $$Q\{Z^{(j)} = z^{(j)}, j=1, \ldots, N\} = \prod_{j=1}^{N} \prod_{i=1}^{k} q_{ij}^{z_i^{(j)}}$$

where

(3.7) $$q_{ij} = p_i e^{ha_j\Delta_i} [\sum_{i=1}^{k} p_i e^{ha_j\Delta_i}]^{-1}.$$

Observe that, under Q, $Z^{(1)}, \ldots, Z^{(N)}$ are independent random vectors taking values in $\Psi$, but are not identically distributed.

__Lemma 3.1__. Let $S_N$ be given by (3.[illegible]), $T_N$ by (3.4), and $P_N(x_N)$ by (2.3). For any $h > 0$ and any $p_1, \ldots, p_k$ with $p_i > 0$, $i = 1, \ldots, k$, $p_1 + \ldots + p_k = 1$, we have

$$(3.8) \qquad P_N(x_N) = \frac{Q\{Z(N) = \underline{n}\}}{P\{Z(N) = \underline{n}\}} \prod_{j=1}^{N} \left[ \sum_{i=1}^{k} p_i e^{h a_j \Delta_i} \right] \cdot A_N,$$

where

$$(3.9) \qquad A_N = \Sigma_{x_N} e^{-ht_N} Q\{T_N = t_N | Z(N) = \underline{n}\},$$

and $\Sigma_{x_N}$ denotes summation over those $t_N$ satisfying $t_N \ge x_N \sigma_\Delta N^{1/2}$.

__Proof__. From (3.6) and (3.7),

$$(3.10) \qquad P_N(x_N) = [P\{Z(N)=\underline{n}\}]^{-1} \Sigma_{x_N} P\{T_N = t_N, Z(N)=\underline{n}\}$$

$$= \frac{\prod_{j=1}^{N} \left[ \sum_{i=1}^{k} p_i e^{h a_j \Delta_i} \right]}{P\{Z(N)=\underline{n}\}} \Sigma_{x_N} e^{-ht_N} Q\{T_N = t_N, Z(N) = \underline{n}\},$$

which is equal to the right hand side of (3.8). Lemma 3.1 is now proved.

Observe that (3.8) is an identity in the k+1 arbitrary variables $p_1, \ldots, p_k$ and h. We will later exploit this fact by making convenient choices for these quantities. But first we will obtain asymptotic approximations to all the factors in the right-hand side of (3.8) except $A_N$. We begin with $P\{Z(N) = \underline{n}\}$.

__Lemma 3.2__. For any $p_1, \ldots, p_k$ such that $p_i > 0$, $i = 1, \ldots, k$ and $p_1 + \ldots + p_k = 1$, we have, under (2.2), as $N \to \infty$,

$$(3.11)\quad P\{Z(N)=\underline{n}\} \sim (2\pi N)^{(1-k)/2} \prod_{i=1}^{k} \lambda_i^{-\frac{1}{2}} \exp[-N \sum_{i=1}^{k} \lambda_i \log(\lambda_i/p_i)].$$

__Proof__. Apply Stirling's formula.

__Lemma 3.3__. Let $q_{i.} = (q_{i1} + \ldots + q_{iN})/N$ and $\underline{\Sigma}_j = \{\sigma_{i\ell}^{(j)}, i, \ell = 1, \ldots, k-1\}$, where

$$\sigma_{i\ell}^{(j)} = q_{ij}(1-q_{ij}) \quad \text{if } i = \ell$$

$$= -q_{ij}q_{\ell j} \quad \text{if } i \neq \ell.$$

Assume $\lim_{N\to\infty} [\sum_{j=1}^{N} (N^{-1}\underline{\Sigma}_j)] = \underline{\Sigma}$ (say) is a positive definite matrix. Then we have, uniformly in $\underline{n}$,

$$(3.12)\quad (2\pi N)^{(k-1)/2}(\det \underline{\Sigma})^{1/2} Q\{Z(N)=\underline{n}\} - \exp\{-\tfrac{1}{2}\, \underline{\xi}_N' \underline{\Sigma}^{-1} \underline{\xi}_N\} \to 0,\ N \to \infty,$$

where $\xi_N' = \{N^{-\frac{1}{2}}(n_1 - Nq_{1.}), \ldots, N^{-\frac{1}{2}}(n_{k-1} - Nq_{k-1.})\}$.

__Proof__. Let $\Xi_N = \{N^{-\frac{1}{2}}[Z_1(N) - Nq_{1.}], \ldots, N^{-\frac{1}{2}}[Z_{k-1}(N) - Nq_{k-1.}]\}$. Clearly $\Xi_N$ is asymptotically normal with zero mean vector and covariance matrix $\underline{\Sigma}$. Rvačeva (1954) has proven a local limit theorem for sums of i.i.d. lattice vectors. To prove (3.12) we follow her argument in outline, varying the details to handle our non-identically distributed vectors and taking advantage of the special structure of $Z(N)$. See Clickner (1972) for the details.

We no. need some notation. Let $E_Q$, $\mathrm{var}_Q$, etc., denote expectation, variance, etc. under Q. Then

$$(3.13)\qquad E_Q T_N = \sum_{j=1}^{N} a_j \sum_{i=1}^{k} \Delta_i q_{ij}$$

and

$$(3.14)\qquad V_T^2 = \mathrm{var}_Q T_N = \sum_{j=1}^{N} a_j^2 \sum_{i=1}^{k} \Delta_i^2 q_{ij}(1-q_{ij}).$$

Also, define $\nu_\Delta = p_1\Delta_1 + \ldots + p_k\Delta_k$ and

$$(3.15)\qquad \tau_\Delta^2 = \sum_{i=1}^{k} p_i(\Delta_i - \nu_\Delta)^2.$$

The motivation for the bound on h in Lemma 3.4 will be made clear in (3.24) and (3.25).

Lemma 3.4. Let $h = h_N$, where $h_N \le Kx_N/N^{1/2}$, for some K, $0 < K < \infty$. Assume $x_N \to \infty$, $x_N^2/N \to 0$, and (2.4) obtains. Then, as $N \to \infty$,

$$(3.16)\qquad \sum_{j=1}^{N} \log\left(\sum_{i=1}^{k} p_i e^{h_N a_j \Delta_i}\right) = Nh_N^2\tau_\Delta^2/2 + O(Nh_N^3\gamma_N^3),$$

$$(3.17)\qquad q_{i.} = p_i + \tfrac{1}{2}p_i h_N^2[(\Delta_i - \nu_\Delta)^2 - \tau_\Delta^2] + O(h_N^3\gamma_N^3),\quad i=1, \ldots, k,$$

$$(3.18)\qquad E_Q T_N = Nh_N\tau_\Delta^2 + O(Nh_N^2\gamma_N^3),$$

$$(3.19)\qquad V_T^2 = N\sum_{i=1}^{k} \Delta_i^2 p_i(1-p_i) + O(Nh_N\gamma_N^3),$$

$$(3.20)\qquad \sum_{j=1}^{N} q_{ij}(1-q_{ij})/N = p_i(1-p_i) + O(h_N^2),\ i=1,\ \ldots,\ k,$$

$$(3.21)\qquad \sum_{j=1}^{N} q_{ij}q_{\ell j}/N = p_i p_\ell + O(h_N^2)\quad,\quad i \neq \ell.$$

Proof. Consider the left-hand side of (3.16):

$$\sum_{i=1}^{k} p_i e^{h_N a_j \Delta_i} = 1 + h_N a_j \nu_\Delta + \tfrac{1}{2} h_N^2 a_j^2 \sum_{i=1}^{k} p_i \Delta_i^2$$

$$+ \frac{\theta}{6} h_N^3 |a_j|^3 \sum_{i=1}^{k} |\Delta_i|^3 e^{h_N |a_j \Delta_i|},$$

where $|\theta| \le 1$. Then

$$\sum_{j=1}^{N} \log\Big(\sum_{i=1}^{k} p_i e^{h_N a_j \Delta_i}\Big) = \sum_{j=1}^{N} \{h_N a_j \nu_\Delta + \tfrac{1}{2} h_N^2 a_j^2 T_\Delta^2 + O(h_N^3 |a_j|^3)\}$$

$$= \tfrac{1}{2} h_N^2 N T_\Delta^2 + O\Big(h_N^3 \sum_{j=1}^{N} |a_j|^3\Big)$$

$$= N h_N^2 T_\Delta^2/2 + O(N h_N^3 \gamma_N^3).$$

Expressions (3.17) - (3.21) follow similarly. The proof of Lemma 3.4 is complete.

Proof of Theorem 2.1

In addition to Lemmas 3.1-3.4, we need two lemmas specifically for Theorem 2.1.

Lemma 3.5. Define $A_N$ by (3.9). Then for each N and any $h > 0$,

$$(3.22)\qquad \begin{aligned} 0 &\geq \log A_N + hx_N\sigma_\Delta N^{\frac{1}{2}} \\ &\geq -4V_T h + \log\{1 - E_Q[(T_N - 2V_T - x_N\sigma_\Delta N^{\frac{1}{2}})^2 | Z(N)=\underline{n}]/4V_T^2\}. \end{aligned}$$

Proof. Clearly $A_N \leq \exp[-hx_N\sigma_\Delta N^{\frac{1}{2}}]$. On the other hand,

$$\begin{aligned} A_N &\geq \exp[-hx_N\sigma_\Delta N^{\frac{1}{2}} - 4V_T h]Q\{x_N\sigma_\Delta N^{\frac{1}{2}} \leq T_N \leq x_N\sigma_\Delta N^{\frac{1}{2}} + 4V_T | Z(N)=\underline{n}\} \\ &\geq \exp[-hx_N\sigma_\Delta N^{\frac{1}{2}} - 4V_T h]\{1 - E_Q[(T_N - x_N\sigma_\Delta N^{\frac{1}{2}} - 2V_T)^2 | Z(N)=\underline{n}]/4V_T^2\}, \end{aligned}$$

by Chebyshev's inequality. Lemma 3.5 follows.

We will now choose values for $p_1, \ldots, p_k$ and h for the k-sample case. Let

$$(3.23)\qquad p_i = \lambda_i, \quad i = 1, \ldots, k$$

and let $h = h_N$ be the unique solution of

$$(3.24)\qquad E_Q T_N = x_N\sigma_\Delta N^{\frac{1}{2}} + 2V_T.$$

This choice of $h_N$ maximizes a term in the lower bound of $A_N$ in Lemma 3.5. A simple arguement shows that $h_N$ is well defined by (3.24), and, further

(3.25) $h_N \sim x_N/\sigma_\Delta N^{1/2}$.

Lemma 3.6. Let (3.23) hold and let $h_N$ be the solution of (3.24), satisfying (3.25). Then, as $N \to \infty$,

(3.26) $E_Q(T_N | Z(N) = \underline{n}) \sim E_Q T_N$

and

(3.27) $\mathrm{var}_Q(T_N | Z(N) = \underline{n}) \sim V_T^2$.

Proof. Consider, for $i = 1, \ldots, k$, $j = 1, \ldots, N$,

(3.28) $Q(Z_i^{(j)} = 1 | Z(N) = \underline{n}) = q_{ij} \dfrac{Q(Z(N) = \underline{n}_i)}{Q(Z(N) = \underline{n})}$,

where $\underline{n}_i = (n_1, \ldots, n_{i-1}, n_i - 1, n_{i+1}, \ldots, n_k)$, $i = 1, \ldots, k$. From (3.17), (3.20), (3.21) and (3.23), $\underline{\Sigma} = \lim N^{-1} \sum_{j=1}^{N} \Sigma_j$ has elements of the form $\lambda_i(1-\lambda_i)$ on the main diagonal and $-\lambda_i\lambda_\ell$ off the main diagonal. Since $\sum_{i=1}^{k} \lambda_i = 1$, $\det \underline{\Sigma} = \prod_{i=1}^{k} \lambda_i$. Hence, by Lemma 3.3,

$$(3.29)\quad \frac{Q(Z(N) = \underline{n}_i)}{Q(Z(N) = \underline{n})} \sim \exp\{\frac{N}{2\lambda_i} [(\lambda_i - q_{i.})^2 - (\lambda_i - 1/N \cdot q_{i.})^2]\}$$

$$= \exp\{(\lambda_i - q_{i.})/\lambda_i \; - 1/2N\lambda_i\}$$

$$= \exp(O(x_N^2/N)),$$

uniformly in $i = 1, \ldots, k$, $j = 1, \ldots, N$. Hence,

$$(3.30)\quad E_Q(T_N|Z(N)=\underline{n}) = \sum_{j=1}^{N} a_j \sum_{i=1}^{k} \Delta_i Q\{Z_i^{(j)} = 1|Z(N)=\underline{n}\} \sim E_Q T_N.$$

Similarly, for $i, i' = 1, \ldots, k$, $j, j' = 1, \ldots, N$, $j \neq j'$,

$$(3.31)\quad Q\{Z_i^{(j)}=1,\ Z_{i'}^{(j')}=1|Z(N)=\underline{n}\} \sim q_{ij} q_{i'j'},$$

uniformly in $i$, $i'$, $j$ and $j'$. The joint probability on the left-hand side of (3.31) is zero if $j = j'$. Hence

$$(3.32)\quad E_Q(T_N^2|Z(N)=\underline{n}) = \sum_{j=1}^{N} a_j^2 \sum_{i=1}^{k} \Delta_i^2 Q\{Z_i^{(j)}=1|Z(N)=\underline{n}\}$$

$$+ \sum_{j\neq j' \text{ or } i\neq i'}\sum\sum\sum a_j a_{j'} \Delta_i \Delta_{i'} Q\{Z_i^{(j)}=1,\ Z_{i'}^{(j')}=1|Z(N)=\underline{n}\} \sim E_Q T_N^2.$$

Lemma 3.6 follows.

We can now prove Theorem 2.1. This will be done by substituting Lemmas 3.2-3.6 in Lemma 3.1. More specifically, let (3.23) hold and let $h = h_N$ solve (3.24) which entails $h_N \sim x_N/N^{\frac{1}{2}}\sigma_\Lambda$. From Lemmas 3.2 and 3.3,

$$(3.33)\quad \frac{Q(Z(N) = \underline{n})}{P(Z(N) = \underline{n})} \sim \exp[-\frac{N}{2}\sum_{i=1}^{k}(\lambda_i - q_{i.})^2/\lambda_i\ ]$$

$$= \exp[O(x_N^4/N)],$$

using (3.17). From Lemma 3.4, specifically (3.16), we have

$$(3.34)\quad \sum_{j=1}^{N}\log(\sum_{i=1}^{k}\lambda_i e^{h_N a_j \Delta_i}) \sim x_N^2/2.$$

By the definition of $h_N$,

$$E_Q[(T_N - x_N\sigma_\Delta N^{\frac{1}{2}} - 2V_T)^2 | Z(N) = \underline{n}] = \mathrm{var}_Q(T_N | Z(N) = \underline{n}) \sim V_T^2,$$

by Lemma 3.6. Hence, from Lemma 3.5,

$$(3.35)\quad \log A_N \sim -x_N^2.$$

Theorem 2.1 now follows by substituting (3.33), (3.34), and (3.35) in Lemma 3.1.

Proof of Theorem 2.2

The main difference between the proofs of Theorems 2.1 and 2.2 lies in the treatment of the sum $A_N$, defined in (3.9). Here,

we write $A_N$ as $A_N = \exp[-hE_QT_N]C_N$,

where

(3.36) $$C_N = \int_B^\infty e^{-hVy}\, d\Phi_N(y),$$

where

(3.37) $$\bar{a} = \sum_{j=1}^N a_j q_{1j}(1-q_{1j}) / \sum_{j=1}^N q_{1j}(1-q_{1j}),$$

(3.38) $$V^2 = \sum_{j=1}^N (a_j - \bar{a})^2 q_{1j}(1-q_{1j}),$$

(3.39) $$\Phi_N(y) = Q\{T_N \le yV + E_QT_N \mid Z(N)=\underline{n}\},$$

and

(3.40) $$B = [x_N(\lambda\bar{\lambda}_N N)^{\frac{1}{2}} - E_QT_N]/V.$$

Observe that $\Phi_N(\cdot)$ is the conditional distribution referred to in the discussion following Theorem 2.2 (see equation (2.14) ).

Now select $p_1$ and h. Choose $p_1 = p_N$ to be the solution of

$$(3.41)\quad n = \sum_{i=1}^{N} \{1+(q_N/p_N)\exp[-x_N a_i(\lambda_N\lambda_N/N)^{1/2}/p_N q_N]\}^{-1}.$$

This is the $p_N$ of (2.13). Choose

$$(3.42)\quad h = h_N = x_N(\lambda_N\bar{\lambda}_N/N)^{1/2}/p_N q_N.$$

Note that $p_N$ is well-defined by (3.41) and we have $p_N \sim \lambda_N$ and $h_N \sim x_N/(\lambda_N\bar{\lambda}_N N)^{1/2}$, so these choices are only slightly different from those in the proof of Theorem 2.1 -- but are more convenient for the proof of Theorem 2.2. The reasons for these choices will become apparent in Lemma 3.7 and display (3.45).

Lemmas 3.7 and 3.8 constitute the analysis of $A_N$.

<u>Lemma 3.7</u>. (Hájek (1964)). Let (3.41) hold. A necessary and sufficient condition for $\Delta_N \to 0$, where

$$\Delta_N = \sup_x |\phi_N(y) - \phi(y)|$$

and $\phi_N(\cdot)$ is defined in (3.39) is

$$(3.43)\quad V^{-2}\sum_{i\in A_\varepsilon}(a_i-\bar{a})^2\pi_i\bar{\pi}_i \to 0$$

for all $\varepsilon > 0$, where $A_\varepsilon = \{i: |a_i-\bar{a}| > \varepsilon V\}$.

<u>Proof</u>. This is Theorem 7.1 of Hájek (1964).

<u>Lemma 3.8.</u> Let $p_N$ and $h_N$ be given by (3.41) and (3.42), respectively, and let (2.9) obtain. With $C_N$ as in (3.36), we have, as $N \to \infty$, $C_N \to 0$.

<u>Proof</u>. We can write

$$(3.44)\quad C_N = \int_B^\infty e^{-h_N yV}\,d\Phi(y)$$

$$+ e^{-h_N VB}[\Phi_N(B)-\Phi(B)] + hV\int_B^\infty e^{-h_N yV}[\Phi_N(y)-\Phi(y)]dy.$$

Techniques similar to those of Lemma 3.4 yield $V^2 \sim N\lambda_N\bar{\lambda}_N$. Consider

$$(3.45)\quad B = V^{-1}[x_N(\lambda_N\bar{\lambda}_N N)^{\frac{1}{2}}-E_Q(T_N)]$$

$$= x_N - pqh(N/\lambda_N\bar{\lambda}_N)^{\frac{1}{2}} + o[(Nh^2)^{-\frac{1}{2}}]$$

$$= o(x_N^{-1}),$$

with h given by (3.41). Hence

(3.46) $C_N = [1+o(1)][(2\pi x_N^2)^{-1/2} + \Delta_N]$.

It follows from (2.9) that $A_\varepsilon$ is empty for large N; hence $\Delta_N \to 0$ by Lemma 3.7. Since $x_N \to \infty$, Lemma 3.8 follows.

To complete the proof of Theorem 2.2, apply condition (2.11) and the selections of $p_1$ and h, (3.42) and (3.43), to Lemmas 3.1 - 3.4 with k=2 to obtain

(3.47) $P_N(x_N) \sim C_N \exp[-x_N^2/2 + J(x_N, \lambda_N)]$.

Proof of Lemmas 2.3 and 2.4

Lemma 2.4 is proved by expanding the logarithms.

In Lemma 2.3, the case $\lambda_N = \frac{1}{2} + O(N^{-1})$ is proved by substituting $p_N = \frac{1}{2} + O(N^{-1})$ to verify that it is a solution of (2.13). Otherwise, recall that

(3.48) $\lambda_N = p + (p-\frac{1}{2})\lambda_N\bar{\lambda}_N x_N^2/pqN + o(N^{-1})$,

using (3.17) and (3.42). Now, suppose $x_N^4 = o(N)$ and propose

(3.49) $p = \lambda_N + c_1 x_N^2/N + o(N^{-1})$

as a solution of (3.48). Substitute (3.49) in (3.48) and solve for

$c_1$ to obtain $c_1 = c_1(\lambda_N)$ with $\lambda_N$, as in (2.20). Next, suppose $x_N^3 = o(N)$ and try

$$(3.50) \quad p = \lambda_N + c_1(\lambda_N) x_N^2/N + c_2 x_N^4/N^2 + o(N^{-1})$$

as a solution of (3.48). Again, solve for $c_2$ to obtain $c_2 = c_2(\lambda_N)$, as in (2.21). The higher order coefficients $\{c_k(\lambda_N)\}$ can be found successively by continuing this iterative procedure. This is not done here because the algebra becomes very cumbersome. The proof of Lemma 2.3 is complete.

Proof of Corollaries 2.5 and 2.6

Corollary 2.5 is an immediate consequence of Lemmas 2.3 and 2.4, and Corollary 2.6 is proved in essentially the same manner as Lemma 2.3.

REFERENCES


[1] Bahadur, R. R. (1960). Stochastic comparison of tests. Ann. Math. Statist. 31 276-295.

[2] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on sums of observations. Ann. Math. Statist. 23 493-507.

[3] Clickner, R.P. (1972). Contribution to rates of convergence and efficiencies in nonparametric statistics. Unpublished dissertation. Florida State University.

[4] Cramér, H. (1938). Sur un nouveau theoreme-limite de la theorie des probabilities. Actualites Sci. Ind. No. 736, 5-23.

[5] Feller, W. (1943). Generalization of a probability limit theorem of Cramér. Trans. Amer. Math. Soc. 54 361-372.

[6] Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist. 35 1491-1521.

[7] Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.

[8] Linnik, Yu. V. (1960). On the probability of large deviations for sums of independent variables. Proc. Fourth Berkeley Symposium in Math. Stat. and Prob. 2 289-306.

[9] Linnik, Yu. V. (1961). Limit theorems for sums of independent variables taking into account large deviations I. Theor. Prob. and its Appl. 6 131-148.

[10] Linnik, Yu. V. (1962). Limit theorems for sums of independent variables taking into account large deviations III. Theor. Prob. and its Appl. 7 115-129.

[11] Rubin, H., and Sethuraman, J. (1965a). Probabilities of moderate deviations. Sankhyā Series A 27 325-346.

[12] Rubin, H., and Sethuraman, J. (1965b). Bayes risk efficiencies. Sankhyā Series A 27 347-356.

[13] Rvačeva, E. L. (1954). On domains of attraction of multi-dimensional distributions (Russian). L'vov. Gos. Univ. Uč. Zap. 29 Ser. Meh. - Mat. No. 6, 5-44. English translation: Selected Transl. Math. Statist. and Probability 2 (1962) 183-206.

[14] Stone, M. (1967). Extreme tail probabilities for the two-sample Wilcoxon statistic. Biometrika 54 629-640.

[15] Stone, M. (1968). Extreme tail probabilities for sampling without replacement and exact Bahadur efficiency of the two-sample normal scores test. Biometrika 55 371-375.

[16] Stone, M. (1969). Approximations to extreme tail probabilities for sampling without replacement. Proc. Camb. Phil. Soc. 66 587-606.

[17] Woodworth, G. (1970). Large deviations and Bahadur efficiency of linear rank statistics. Ann. Math. Statist. [illegible]